**Purpose**: follow this checklist to help determine a good target for our virtual drug screening research.

Do not start this protocol until you have finished your VS3 run2.

In short, we want to find an enzyme that is crucial for the function or survival of a pathogenic organism. We need to have some confidence that the enzyme could be expressed and purified in the wet lab and we would like for the activity of the enzyme to be testable through spectrophotometric methods. It is best if the reagents for the enzyme assay are neither overly expensive nor difficult to obtain. Another requirement is the ability to carry out virtual screening on the enzyme target. Either the crystal structure of the enzyme or the structure of a homologous protein that is similar enough to where we could make a homology model is necessary to make virtual screening possible.

You could start with the protein target from VS3 that you chose. However, you will need to determine essentiality of that protein before it is deemed suitable. Also, if it is an old entry – that probably means a lot of work has already been done on that target. If you really want a novel target – you should start from scratch.

Once you have found a target that seems like it is promising – create a Wikispaces ‘**Target Page’**. There are instructions for how to create a page on the main Targets page and a list of the criteria to enter for your target.

All of the information found, should go in your notebook too – even though we may not keep the original target that you choose.

Don’t choose a target that has been taken by a 2012 VDSer or a 2013 VDSer (check the Wikispaces Targets Page). However, you can choose a target that is in the same organism as another researcher’s target (e.g. *Yersinia pestis* and *Yersinia pestis*)

Note that this is an **iterative** process. You will probably start on something, realize it isn’t good and restart. Or Dr .B and the mentors will shoot down your target for some other reason! Don’t get frustrated – target selection is a frustrating process. Welcome to research!

After everyone has posted their targets to Wikispaces, we will review them and select the best ones (4 or 5) to move forward.

**CHOOSE ORGANISM**

Requirements: should be bacterial or protozoan (maybe fungal or viral if compelling case)

Be sure to determine which strain or serotype it is. The organism needs to be responsible for a disease (either a regular infectious or a potential bioterrorist weapon disease).

NIAID Emerging and Re-emerging Infectious Diseases – Group I, II, and III

<http://www.niaid.nih.gov/topics/emerging/pages/list.aspx>

NIAID Biodefense – Group III: Category A, B, and C Priority Pathogens

<http://www.niaid.nih.gov/topics/BiodefenseRelated/Biodefense/research/Pages/CatA.aspx>

For the following databases, it is usually best to start with the EC number and then search from there

E.C. number (e.g. 3.\* find hydrolases)

1. TDR targets webpage [www.tdrtargets.org](http://www.tdrtargets.org/)
2. J Craig Venter Institute – New Infectious Diseases Targets <http://gsc.jcvi.org/projects.html>
3. Neglected and Tropical Diseases (PLoS One) - <http://www.plosntds.org/static/scope.action>

They have about 10 bacterial targets

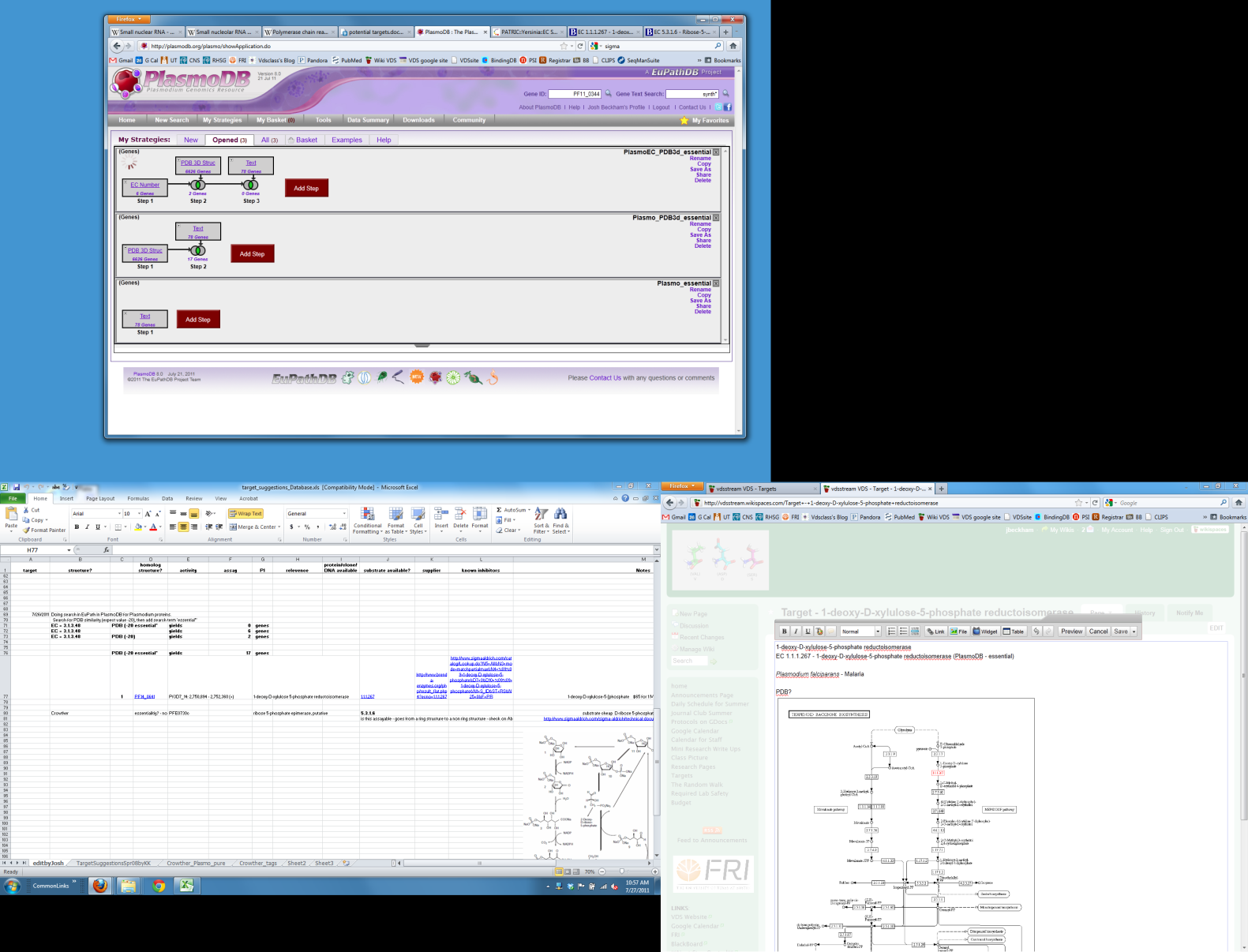
1. BRC (Bioinformatics Resource Centers) from NIAID

<http://www.niaid.nih.gov/LabsAndResources/resources/dmid/brc/Pages/awards.aspx>

1. COMBINED PORTAL: <http://pathogenportal.net/portal/portal/PathPort/Home>
2. **EUPATH**: Eukaryotic Pathogens Database Resources: <http://eupathdb.org/eupathdb/>
   * Either click on one of the 9 organisms present to do a specific search within that database – or use the main page to search all of the databases.

Amoeba, Giardia, Cryptosporidium, Microsporidia, Piropolasma, Plasmodium, Toxoplasmosis, Trichomonas (unsupported), TriTryp (Trypanosoma, Leishmania)

* + See the ‘About’ tab and ‘Organism in Database’ to see what is there



**Fig. 1.** Example of multiple searches with multiple parameters in PlasmoDB

1. PATRIC: **Bacteria**: <https://patricbrc.vbi.vt.edu/portal/portal/patric/Home>
   * + - Confirm ‘security exception for this site’
       - Click the **ORGANISMS** tab
       - Then click on **EC Search**
     + Change **keyword** to **EC number** and enter the number you want
     + or search for EC Number under SEARCHES & TOOLS tab
       - use **Accession Number** or **Locus Tag** to see results
2. ViPR: **Viruses**: <http://www.viprbrc.org>

Etiologic Risk Group Categories (for pathogens):

<http://www.utexas.edu/research/rsc/ibc/agent_class.html#_Toc7238334>

Record background information on the disease that could be used in the Intro of a paper.

Does the enzyme work in a complex of other proteins/macromolecules or does it do its job alone?

**ESSENTIALITY INFO**

Ideally we want an enzyme that is crucial for the survival or virulence of the organism. Some of the databases have information on this however, it may be necessary to read into the journal papers some to determine how important the target is. Often essentiality info is not for the organism you are looking at but rather for another organism for which experimental evidence exists. If there is essentiality info – you will want to expand the list and determine from which organism the information came. If it was a study on blue whales and you are studying malaria parasite – that essentiality info is not very useful!

NOTE: Can add a text search to EuPath (and related databases) or PATRIC for ‘essential\*’

NOTE: for Eukaryotic pathogens – the enzyme may only be important in a specific life cycle stage. Make note of this too

NOTE: In TDR Targets you can search by keyword ‘**carbonic’ or ‘phosphatase’** or ‘**reductase’** and then specify that the results need to have **Essentiality**

**DRUGGABLE TARGET**

In the databases – determine if the protein is a ‘druggable’ one – i.e. is it similar to other known targets that have been successfully inhibited/bound by current drugs

This info may or may not be present. You may need to read into the journal papers for the protein.

**TDRTargets** has a ‘Druggability Index’ for its entries.

EXTRAS:

Video to see how to do a search for a druggable protein target (Part 3 of video set)

<http://www.ibioseminars.org/lectures/global-health-a-energy/david-s-roos/david-s-roos-part-3c.html>

Target Discovery Tutorial website (covers more detail than we need for our purposes – i.e. the second half):

<http://www.alzforum.org/drg/tut/tutorial.asp>

**ASSAYABLE ENZYME TARGET**

Requirements:

1. Need to determine if the enzyme is assayable

Usually want a light based assay (absorbance or fluorescence)

1. Substrate needs to be readily available

e.g. 4-nitrophenyl acetate (PNPA) or 4-nitrophenyl phosphate (PNPP), or NADPH, NADH

BRENDA webpage for enzymes – can search by substrate by doing advanced search

<http://www.brenda-enzymes.org/>

TDR site links you to Brenda enzyme assay for each one. [www.tdrtargets.org](http://www.tdrtargets.org/)

SIGMA - Or any other chemical supplier

see if there is an enzyme assay protocol for it by searching under the EC number <http://www.sigmaaldrich.com/life-science/metabolomics/enzyme-explorer/learning-center/assay-library/ec-number.html>

Easy/cheap enzyme assays – phosphatases, carbonic anhydrases, reductases etc.

NOTE: can search PATRIC and EuPath by E.C. number

Phosphatases (EC # 3.1.3\*.., EC# 3.1.3.2, EC# 3.1.3.41), tyrosine phosphatases EC# 3.1.3.48

Carbonic anhydrases (EC # 4.2.1.1), Carboxylesterase (EC # 3.1.1.1), Acetylesterase (EC # 3.1.1.6). Reductases (EC # 1.\*)

ExPASy Proteomis Server – search by E.C. <http://ca.expasy.org/enzyme/>

After searching, click on:

“[All UniProtKB/Swiss-Prot entries referenced in this entry](http://www.uniprot.org/uniprot/?query=ec:3.1.1.1%20reviewed:yes), with possibility to download in different formats, align etc.”

As an alternative - can do a binding assay instead of an enzyme assay.

Eg. Differential Scanning Fluorimetry (DSF)

Determine cost of reagents e.g. substrates, cofactors (usually from Sigma). Often find the smallest quantity available and use that price.

**STRUCTURE AVAILABLE**

Basically you want to see if your protein might be similar to something in the PDB so that you will have a crystal structure on which to do virtual screening. All of the proteins that were chosen in VS3 have crystal structures since they are in the PDB. However, if you would like to ‘cast a wider net’ and explore other targets or even organisms that are not yet in the PDB – you can make a homology model later on to do virtual screening.

1. If you don’t have a particular protein yet
   * Check PDB for Xray crystal structure <http://www.pdb.org>

Can search PDB by EC# specifically (advanced search)

1. EuPath has a search feature to determine sequence homology to PDB structure.

* As an example, do an E.C. number search in EuPath or PATRIC.
* Select your organism (species)
* New Search >> Search for Genes >> Putative Function >> E.C. Number
* Then, you can add search feature that specifies predicted similarity of your target to any protein in the PDB database.
  + Add Step >> Protein Attributes >> Predicted 3D Structure >>
  + With similarity to PDB Protein from: ‘**select all’**
  + Set ***BLAST P-value less than 10 to the*** : -20 (i.e. x10-20).

The **E value**, or **Expect value**, is a parameter that describes the number of hits one can expect to see just by chance when searching a database of a particular size. For example, an E value of one indicates that a result will contain one sequence with similar score simply by chance. The scoring takes chain length into consideration and therefore shorter sequences can have identical matches with high E value.

OTHER OPTIONAL SITES – that may also have a crystal structure

Check PSI (Protein Structure Initiative) <http://sbkb.org/index.html>

Check SGC (Structural Genomics Consortium) <http://www.thesgc.org/>

Check Center for Structural Genomics of Infectious Diseases (CSGID) <http://csgid.org/csgid/cake/targets/index>

If **no** Xray crystal structure – then could you make a homology model?

**IF**: Amino acid sequence is **> 50%** similar to enzyme of the same functionality in the PDB

* + Get the amino acid sequence for your protein (from NCBI, ExPASy, PATRIC, EuPATH, JCVI or etc.) Use NCBI BLASTP search and select PDB <http://blast.ncbi.nlm.nih.gov/Blast.cgi>
* Note: searching this way is basically making a Pairwise comparison - sort of like you did in the VS2 lab – but it is comparing to ALL of the proteins in the PDB.
* Look at the Max Ident and the % Positives
  + The %Positives is what needs to be at least 50% in order to make a homology model
* Also, note which chain of the template protein in the PDB would be used to make the homology model

**THEN**: You can make a Homology Model at a later time

PSI Homology Model portal <http://www.proteinmodelportal.org/>

SWISS-MODEL <http://swissmodel.expasy.org/>

Nature Protocol: <http://www.nature.com/nprot/journal/v4/n1/abs/nprot.2008.197.html>

SCWRL server <http://www1.jcsg.org/scripts/prod/scwrl/serve.cgi>

ICM <http://www.molsoft.com/homology.html>

**SIMILARITY TO HUMAN**

Ideally – if searching for anti-infectives, you don’t want your enzyme to be similar to human form.

Run a BLASTP of your amino acid sequence (protein) against the human database.

Show how similar your protein is to any human proteins by finding where in the list the first human homolog pops up.

**CHECK CURRENT INHIBITORS (is it an over studied target?)**

* Binding DB <http://www.bindingdb.org/bind/index.jsp>
  + Click on “Number of Compounds” under Target. If there are only a few compounds for this target, it may be a good one to study because not many others have looked at it.
* [Drugbank](http://www.drugbank.ca/)
* [PDB](http://www.pdb.org/pdb/home/home.do)
* [BRENDA](http://www.brenda-enzymes.info/)

**EXPRESSION**

Is your gene expressed in the target organism at high levels? Is it only expressed at certain times? What impact may this have upon a drug delivery strategy?

Can it be expressed in bacteria - E. coli BL21(DE3) ? Check [BRENDA](http://www.brenda-enzymes.info/) and literature in [PubMED](http://www.ncbi.nlm.nih.gov/pubmed/)

Human enzymes may be tricky to express in bacteria (glycosylation, post-translational modification

Protein localization - You don’t want a membrane bound protein (unless the part you are interested is outside the membrane) because it won’t be very soluble

TDR Targets and the EuPath databases have options to check for ‘no transmembrane regions’

Check probability of membrane bound regions by running your sequence through the server:

<http://www.ch.embnet.org/software/TMPRED_form.html>

**GENE AVAILABLE**

1. **Plasmid or Genomic DNA Available?**

Check Harvard/ASU PSI

SGC clones through OpenBiosystems (Thermo Fisher) – list in SGC folder

SGC is mainly human structures, some Plasmodium

ATCC – vectors and genomic

<https://www.atcc.org/CulturesandProducts/Microbiology/DNAandRNA/tabid/181/Default.aspx#bacteria>

Also check ATCC under Cell Lines first, then go to the Genomic DNA for that cell line

IMAGE consortium

MGC – mammalian genome collections

1. **Use Synthesis from scratch to Make Any Gene**

**Overlap Extension PCR –** technique stolen from BioBricks

The amino acid sequence could then be codon optimized for expression in *E. coli* bacteria

<http://helixweb.nih.gov/dnaworks/dnaworks_help.html>

PSI uses **GenScript** for gene synthesis – and codon optimization

GENE: To find sequence, go to NCBI gene. Click on the NM link (mRNA) and then find the DNA sequence for the coding region (the CDS) which encodes the amino acids for the protein.

PROTEIN: Also at the bottom of the gene page will be the NP – or protein sequence – show it in FASTA format by clicking at the top left of the Protein page